PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

The Habitable-Zone Planet Finder: improved flux image generation algorithms for H2RG up-the-ramp data

J. P. Ninan, Chad F. Bender, Suvrath Mahadevan, Eric B. Ford, Andrew J. Monson, et al.

J. P. Ninan, Chad F. Bender, Suvrath Mahadevan, Eric B. Ford, Andrew J. Monson, Kyle F. Kaplan, Ryan C. Terrien, Arpita Roy, Paul M. Robertson, Shubham Kanodia, Gudmundur K. Stefansson, "The Habitable-Zone Planet Finder: improved flux image generation algorithms for H2RG up-the-ramp data," Proc. SPIE 10709, High Energy, Optical, and Infrared Detectors for Astronomy VIII, 107092U (2 August 2018); doi: 10.1117/12.2312787

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2018, Austin, Texas, United States

The Habitable-Zone Planet Finder: Improved flux image generation algorithms for H2RG up-the-ramp data

J. P. Ninan[a,b], Chad F. Bender[c], Suvrath Mahadevan[a,b], Eric B. Ford[a,b], Andrew J. Monson[a,b], Kyle F. Kaplan[c], Ryan C. Terrien[d], Arpita Roy[e,f], Paul M. Robertson[g], Shubham Kanodia[a,b], and Gudmundur K. Stefansson[a,b,h]

[a]Department of Astronomy & Astrophysics, The Pennsylvania State University, 525 Davey Lab, University Park, PA, USA, 16802
[b]Center for Exoplanets & Habitable Worlds, The Pennsylvania State University, University Park, PA, USA, 16802
[c]Steward Observatory, University of Arizona, Tucson, Arizona, USA, 85721
[d]Department of Physics and Astronomy, Carleton College, Northfield, MN, USA, 55057
[h]NASA Earth and Space Science Fellow
[e]Milikan Fellow
[f]Department of Astronomy, California Institute of Technology, Pasadena, CA, USA, 91125
[g]Department of Physics & Astronomy, University of California - Irvine, 4129 Frederick Reines Hall, Irvine, CA, USA, 92697

ABSTRACT

Noise and stability of current state of the art near-infrared (NIR) array detectors are still substantially worse than optical science grade CCDs used in astronomy. Obtaining the maximum signal-to-noise ratio in flux image is important for many NIR instruments, as is stable well understood data reduction and extraction. The Habitable-zone Planet Finder (HPF) is a near-infrared ultra stable precision radial velocity (RV) spectrograph commissioned on 10-m Hobby-Eberly Telescope (HET), McDonald Observatory, Texas, USA. HPF uses a Teledyne H2RG array detector. In order to achieve the high-precision ($\sim$ 1 m/s) RV measurements from the NIR spectrum of HPF's science target stars, it is vital to maximize the signal-to-noise ratio and to accurately propagate the uncertainties. Here we present the algorithms we have developed to significantly improve the quality of flux images calculated from the up-the-ramp readout mode of H2RG. The algorithms in the tool `HxRGproc` presented in this manuscript are used for HPF's bias noise removal, non-linearity correction, cosmic ray correction, slope/flux and variance image calculation.

Keywords: H2RG NIR detector, HxRGproc, Up-the-ramp image processing, Algorithms, Precision RV

1. INTRODUCTION

The Habitable-Zone Planet Finder (HPF) is a fiber-fed near-infrared (0.808 to 1.279 μm) ultra-stable high-resolution precision radial velocity (RV) spectrograph recently commissioned on the 10-m Hobby-Eberly Telescope (HET) at McDonald Observatory, Texas, USA.[1–3] HPF uses a 1.7 μm cutoff Teledyne H2RG (Hawaii-2RG HgCdTe 2048x2048) array as the detector. In order to achieve the 1 m/s RV precision goal, we have to detect Doppler shift of $1/2000^{th}$ of an 18 micron H2RG pixel*. Such an RV precision measurement is currently an uncharted regime for NIR detectors. Unlike optical CCDs, images from H2RG arrays have significant non-Gaussian spatial and temporal noise. Raw read-out signals have to be corrected for bias fluctuations, non-linearity, persistence, and other effects.

Further author information: E-mail: jpn23@psu.edu

*This level of precision measurement is realized by effectively measuring shifts from all parts of the spectrum spanning thousands of pixels simultaneously. Any systematic errors from the detector which does not average out will limit achievable RV precision.

High Energy, Optical, and Infrared Detectors for Astronomy VIII, edited by Andrew D. Holland, James Beletic, Proc. of SPIE Vol. 10709, 107092U · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2312787

In this manuscript we describe the reasoning behind the techniques, algorithms, and the software tool (`HxRGproc`) we developed to clean and process H2RG up-the-ramp data. While the algorithms were developed to meet the science goals of HPF spectrograph, they are directly applicable to any HxRG based astronomical instrument. In the subsequent sections we summarize various sub-modules and algorithms of our H2RG data processing module which are used to generate the clean 2D flux images that the spectral extraction algorithms use to extract 1-D spectra.

2. HPF'S H2RG READOUT MODE

The NIR array's readout voltage stability as well as thermal stability are crucial to avoid introducing any systematic artifacts in the RV measurement with HPF. Our thermal control load measurements showed clear transients as we switch between various readout modes of the H2RG (i.e. from 4 to 32 channel, or different reset modes). Hence, during science operation we consistently operate the detector in pixel level reset - 4 channel readout mode. In this mode the detector is continuously reset at 10.45 seconds cycle while idle. Once an exposure is triggered, it will readout the whole array in 4 vertical parallel channels at the same cadence of 10.45 seconds. The non-destructive readout is continued till the end of the exposure, after which the reset cycles restart. For more details on the readout pattern and detector pixels see 4, 5. We do not perform any grouping of readouts in our up-the-ramp. The first non-destructive readout immediately after the reset loop is stopped is called Pedestal frame[†]. Ground based observations through variable weather, guiding errors and the large variability in flux across a single track of the HET (due to the unique design of the telescope) all break assumptions of constant flux. In the rest of the manuscript we describe how this complex and variable time series up-the-ramp 3D data with bias variations, non-linearity, and other artifacts is processed to obtain a single clean 2D flux (slope) image.

3. BIAS FLUCTUATION

The biggest source of non-Gaussian errors in an H2RG based system is the bias fluctuation during the array readout. This is an additive signal, which gets added on top of the pixel readout down the signal chain before digitization in the SIDECAR ASIC (which we use for HPF). This additive nature is verified by taking saturated up-the-ramp data, and measuring the signal fluctuations on top of the steady saturated pixel count signal. This signal is highly correlated and has a 1/f noise spectrum.[6] Figure 1 shows a typical example of bias fluctuation we measured in our detector system. H2RG detectors have four rows or columns of reference pixels all around the edges, which are not light sensitive, for measuring these bias signals to correct them.

We subtract this bias fluctuation before applying any other detector corrections such as non-linearity. Bias fluctuation of variance σ^2_{bias} is an additional noise term in the flux/slope image calculation of the up-the-ramp data. For a single pixel this noise term behaves exactly like readout noise (σ^2_{ron}) or bit quantization error. However, the crucial difference between this bias noise term and others is that it is correlated across pixels in the array. Reducing this error term to less than the per pixel readout noise typically suffices for many imaging applications. However, precision RVs are calculated from the entire spectrum spread over many pixels, and it is sensitive to any systematic noise structure in the image which does not average down like readout noise across pixels. Hence, it is quite important to reduce these highly spatially correlated bias noise to a significantly smaller level than the averaged out readout noise.

Equation 1 is the variance formula for the slope estimate. This formula is obtained by updating the formula derived by Robberto M. 2010[‡], with the σ^2_{bias} term.

$$\sigma_b^2 = \frac{6}{5}\frac{(n^2+1)}{n(n^2-1)}\frac{b}{t_f} + 12\frac{1}{n(n^2-1)}\frac{1}{t_f^2}[\sigma^2_{ron} + \sigma^2_{bias} + \frac{g^2}{12}], \quad (1)$$

[†]This frame is not a zero integration time frame. Since we are doing pixel level reset, the exposure time is same as a single frame readout time.

[‡]JWST-STScI-002161,https://jwst.stsci.edu/files/live/sites/jwst/files/home/instrumentation/technical%20documents/JWST-STScI-002161.pdf

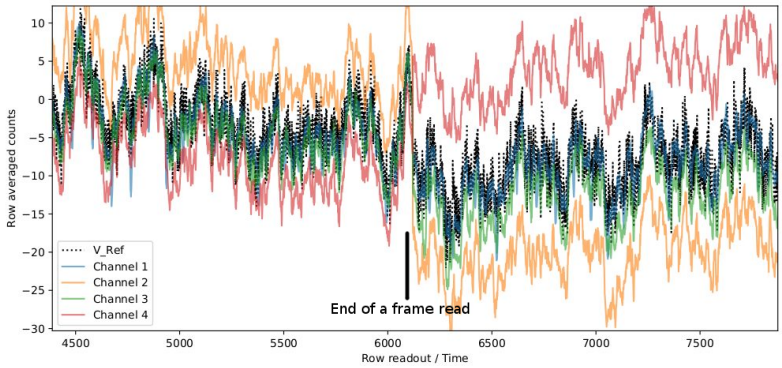

Figure 1. Plot showing the time series fluctuation in the bias value during the read out of H2RG detector. X axis is an arbitrary range of row readout numbers representing time of the readout. Each curve is a row average of separate readout channels (readout noise gets averaged out revealing the underlying bias fluctuations). The Pedestal difference of each channel is subtracted out from this plot for clarity. The sudden change in bias levels across different channels at the end of a frame readout is marked. V_Ref curve with significantly more noise is the average of 4 column vertical reference pixels at the edge of Channel 1.

where σ_b^2 is the variance of the slope estimated from the up-the-ramp data (in units of $(e^-/sec)^2$), n is the number of non-destructive readouts, b is the true flux slope, t_f is the one full frame readout time, σ_{ron}^2 is the readout-noise variance, σ_{bias}^2 is the bias variance (this is correlated across pixels), and g is the e^- to ADU gain.

3.1 Basic Bias Subtraction

During the raster scan readout of each channel of the detector, we get an independent measurement of the bias in each channel while a reference pixel is read out. In the traditional readout mode at which we are operating HPF, we have reference pixels at the first and the last four rows of each channel readout for pinning down the bias differences between the four channels. We also have the first and the last four columns of the first and the last channels as reference pixels. These sample the bias values as each row is read out[§]. The bias signal measured from these reference pixels are conventionally filtered and subtracted to remove the bias fluctuations.[8,9] In this subsection we describe our optimized version of the conventional bias subtraction procedure using the reference pixels.

As shown in Figure 1, there is a significant jump in the bias values of each channel at the end of a frame readout. This sudden offset is slowly varying during the read out of a channel, and the linear component of this variation is subtracted out by a linear interpolation of the bias values at the top and bottom four rows of reference pixels. Since odd and even column pixels have different bias variations (alternative column noise (ACN)), this calculation is done for odd and even columns of the channel separately. In our four channel readout mode, we have a total of 1024 reference pixels for estimation of this linear component. The variance of the most efficient estimate of average due to readout noise is $(\sigma_{ron}/32)^2$. Usage of median for robustness will further increase the variance by a factor of $\pi/2$. Hence, we used a significantly more efficient estimator than median, and at the same time more robust estimator than the mean, viz. "biweight location" to estimate the mean bias value from the reference pixels.[10]

Figure 2 shows the partition of a single non-destructive readout frame for bias correction. While using robust order statistic based estimators like median and biweight location, it is important to remove any systematic offsets between the pixel values (else the variance of the estimate will be as high as σ_{ron}^2). H2RG detectors have

[§]The IRS^2 readout mode proposed by 7 does interleaved reference pixel reads, thereby obtaining a more simultaneous and accurate bias measurement. However, currently, we do not operate in that readout mode since standard SAM + SIDECAR control hardware does not support IRS^2 readout. The IRS^2 is a significant deviation from the single pixel reset cycles which happen before exposure starts. The thermal stability implications of this for RV measurements need to be studied more.

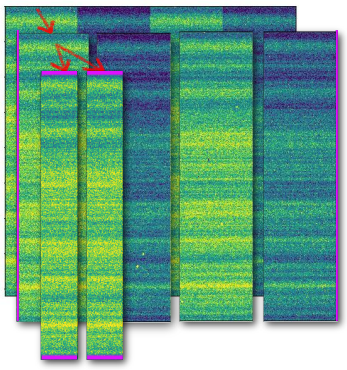

Figure 2. Partitioning diagram of a non-destructive readout frame along the readout channels for estimating the bias from the reference pixels (marked in pink). The reference bias calculation is done separately for odd and even columns of each channel as shown in the diagram. The intensity map shows the bias variations across the channels as well as inside a channel manifesting in horizontal bands.

significant pedestal noise which raises each pixel to a different base level. In order to remove these offsets, we subtract the first Pedestal non-destructive readout frame from all subsequent frames. This process will increase the effective σ_{ron} to $\sqrt{2}\sigma_{ron}$ for reference pixels. However, it does not directly increase the effective readout noise for the final slope image, since it only affects the y intercept of the slope line fit[¶]. This step does not reduce the channel bias fluctuation variance to $(\sqrt{2}\sigma_{ron}/32)^2$ level, due to the inability of the linear interpolation to remove the higher order fluctuations during a readout. It is compounded by the fact that the four rows of the reference pixels only sample an instance of the short period fluctuations shown in Figure 1. This would have been a more significant issue, if 32 channel readout mode of H2RG was used instead of the 4 channel mode.

The next step is the subtraction of the higher frequency fluctuations common to all the four channels (these are the horizontal banding seen in Figure 2). We have independent measurement of this variation from the four reference pixel columns on both sides. If we use mean estimator to average 8 pixels of each row, our estimate will be readout noise limited at $\sqrt{2}\sigma_{ron}/\sqrt{8}$ level. This will result in a significant high frequency horizontal banding in the final slope image. To reduce introduction of this readout noise artifact, we need to smooth the vertical reference pixel measurement. However, this will be at the cost of inability to correct high frequency components of the bias fluctuations. Figure 3 shows the effective noise in the data for a range of window sizes parameter of 2 degree Savitzky-Golay filter. We found that a vertical filter window size of 15 rows was optimal for HPF's H2RG detector.

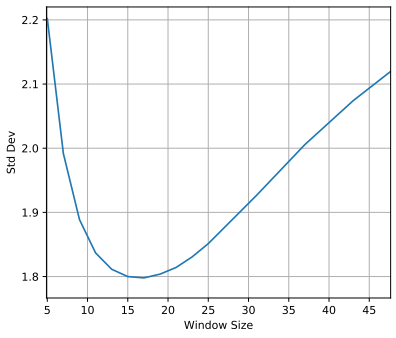

Figure 3. Optimization of a Savitzky-Golay filter with 2 degree polynomial and variable window size for optimal compromise between removal of high frequency vertical bias fluctuations and readout noise artifact after subtraction.

[¶]We will later estimate the lost flux in pedestal subtraction for including in the non-linearity correction calculation.

3.2 Advanced Bias Subtraction

Using just the reference pixels as we described in the previous section will enable reduction of effective σ_{bias} to less than a small fraction of σ_{ron} per pixel. However, since σ_{bias} does not average down like σ_{ron} when using flux from nearby pixels, we will soon get limited by the effective σ_{bias} noise. For example, the green curve in Figure 4 shows the jumps at channel boundaries due to residual σ_{bias}, they are visible in the plot after averaging out readout noise by averaging along the vertical columns. If left uncorrected even these residual fluctuations can impact the precision of the measured radial velocity. HPF's spectral format results in a significant fraction of pixels between echelle orders which are only illuminated by the smooth scattered light. This enables us to exploit the correlated signal information across these pixels to better constrain the bias fluctuations to higher accuracy than what is possible with reference pixels alone.

3.2.1 Median Residue Correction Method

This method is to improve correction for the vertical fluctuations across read out channels. Fundamentally, this method uses the following two assumptions for the low percentile count pixels (mostly inter echelle order pixels) in a channel.
1) Since the flux is smoothly varying across pixels, the mean flux of odd column pixels is the same as mean flux of the even column pixels in a channel during an exposure.
2) The counts in the low flux pixels should linearly increase with time during an up-the-ramp exposure.
A robust estimate of the median pixel count value based on only the low percentile flux pixels is calculated for top and bottom halves of each channel. This average estimate of about 2.6×10^5 pixels has readout noise contribution averaged down to $\sim \sigma_{ron}/500$. A linear slope is fitted to this sequence of up-the-ramp median values to obtain deviations from a straight line. These deviations are mostly the σ_{bias} noise which did not average out across the pixels. This residue is taken as the second order bias correction and applied on to the data. Since this method uses the median of the pixels across the channel instead of four reference rows, it also provides a better averaging over the high frequency horizontal banding noise. The red curve in Figure 4 shows the result for comparison after application of this method, which is particularly well suited to the echelle high resolution spectroscopy format.

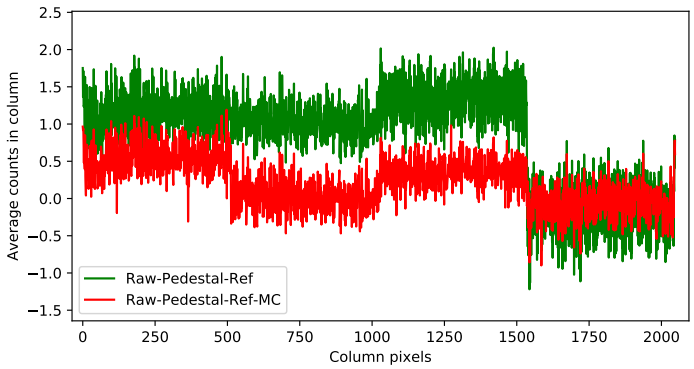

Figure 4. Vertical column averaged signal before (green) and after (red) our median residue correction method is shown. Green curve shows significant jump in counts across four channels, as well as higher noise. The channel offset in green curve is significantly higher than the readout noise limited bias estimate error. Analysis of vertical columns shows that this offset in average is driven by the high frequency fluctuations in the channel bias which are not removed in a simple linear interpolation model. Usage of pixels across the channel in our median residue correction method improves this to a significant extent. The overall high frequency noise across odd and even columns inside each channel is also improved.

3.2.2 Vertical High Frequency Bias Correction

More than the channel bias fluctuations we dealt in previous subsection, the vertical high frequency bias fluctuations are limited by the small number statistics of the reference pixels. The blue and orange curves in Figure 5 show row averaged (to average out readout-noise) plot of a section of two separate channels. The correlated

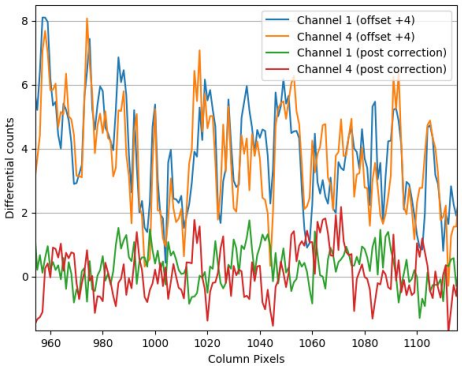

Figure 5. Channel averaged signals before and after our iterative correction is shown. Before the correction one can see a significant correlated noise in two independent channels. The corrected data (red and green) have no significant correlation.

high frequency σ_{bias} is visible in both channels even after the basic reference pixel subtraction. This structure is the cause of the residual faint horizontal banding noise. Since HPF's echelle orders are along the horizontal pixels, we cannot use the previous technique to estimate this signal. Instead, we calculate it iteratively. The first generated slope image is used as a template to subtract the smooth scattered flux in inter echelle order pixels of the up-the-ramp frame. The rows are then weighted averaged to obtain the vertical bias correction values. This is then applied to the frames before the next iteration of slope image calculation. Typically one iteration is enough. The green and red curves in Figure 5 show the non-correlated channel after the application of this iterative technique. Robustness of this method under different flux conditions is still under study, and we have not yet enabled this correction by default to all data in HPF pipeline.

Any remaining bias fluctuation induced slope changes gets further removed significantly during sky subtraction due to the associative property of the slope calculation formula. We discuss more on this property in Section 7.

4. NON-LINEARITY CORRECTION

After bias correction, the next major correction to the up-the-ramp data is the classical non-linearity correction. Unlike CCDs, the H2RG pixels have different non-linearity response across pixels. The non-linearity correction curves have to be calibrated at a pixel by pixel level. The conventional method of fitting an interpolation function to transform the measured counts to the linear extrapolation of the flux estimate from the low ($\sim$5%) well depth counts[8,11] is susceptible to many systematic biases for the following reasons.
1) A significant fraction of pixels have varying levels of non-linearities even at small fractions of the well depth. This will result in systematically under estimating the non-linearity.
2) The first few readouts of the up-the-ramp data is usually affected by the persistence from previous saturation while taking calibration data.
3) A simple lower order polynomial fit typically cannot fit the bottom linear part and highly non-linear top part of the curve simultaneously.
4) This approach is sensitive to model uncertainties and systematic biases introduced due to the choice of threshold which defines the linear range inside the well depth.

Since we know the non-linearity curve is a monotonically decreasing function, a more constrained fit can be made by fitting the change in count versus time (Figure 6). If c is the measured count at time t for a constant flux source F_o, the change in the counts as a function of time is given by the differential equation $\frac{dc}{dt} = F_o\epsilon(c)$. We can model $\epsilon(c)$ with a monotonically decreasing polynomial with slope constrained to be zero at $c = 0$. The fitted function can then be integrated $\int_0^T F_o dt = \int_0^C \frac{dc}{\epsilon(c)}$ to obtain the non-linearity correction function $NLC(C)$ which maps measured counts C to the true count F_oT. Since the divergence of a Taylor series approximation

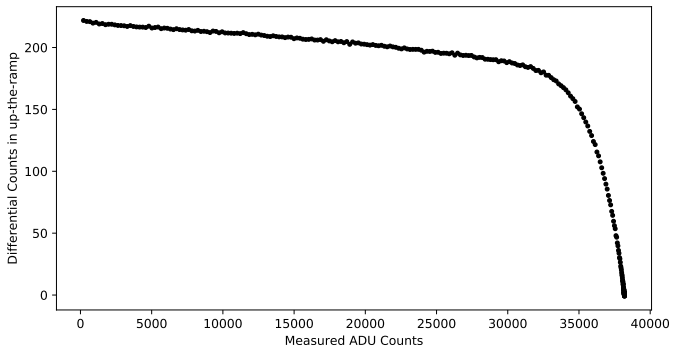

Figure 6. Change in the measured count rate as a function of measured counts during an exposure of a constant flux source. This monotonically decreasing function is modeled as $\epsilon(c)$ in our models. For our Inv B-spline model, inverse of this curve $1/\epsilon(c)$ is modeled using B-splines.

is proportional to the order and distance to the nearest pole, it can be improved by dividing the polynomial by another polynomial which has similar order pole at the same location. This makes the Padè approximation polynomials better for fitting this curve $\epsilon(c)$ which has a pole where c equal to the saturation value of pixel. However, the degree of the polynomial in denominator of Padè approximation could not be made more than 2 to keep the $\frac{dc}{\epsilon(c)}$ analytically integrable. The major motivation for analytical polynomial models for non-linearity curve is to avoid the computational bottleneck of any computationally intensive method for applying corrections on $\sim$ 4 million pixels of H2RG. Since integration of B-splines are computationally fast, we also modeled the inverse of Figure 6 (i.e. $\frac{1}{\epsilon(c)}$) with B-splines. Figure 7 shows the comparison of various best fitted non-linearity correction models. Inv B-splines model was found to be most robust across 4 million pixels. The only draw back of this model is the I/O bottleneck the large coefficient array creates when it is required to be saved or loaded. However, once loaded the computation is fast.

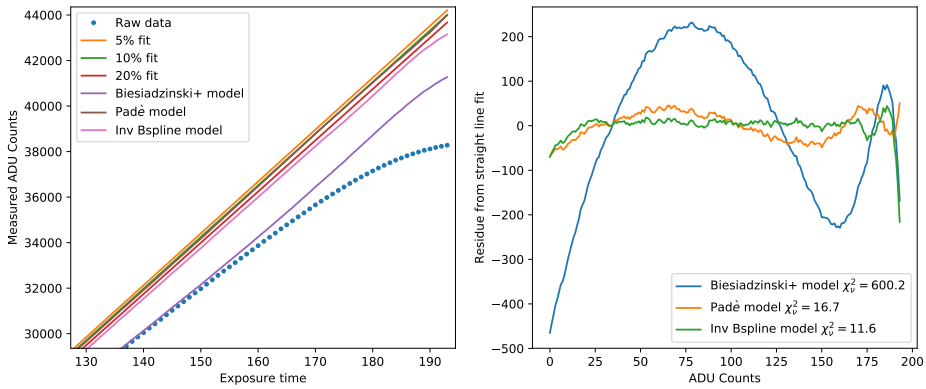

Figure 7. First panel shows the comparison of various non-linearity correction models we developed. The 5%, 10% and 20% fit lines are the linear extrapolation of corresponding well depths. Biesiadzinski+ model is the model proposed by 12, Padè model and Inv Bspline model are described in the text. On the right side is the residue of a stright line fit to the linearity corrected data by the three models. The reduced χ^2_ν estimate showing the goodness of the straight line fit is shown in the legends.

For fitting the non-linearity correction curves we took multiple saturating up-the-ramp data of constant illumination source. The persistence of the detector contaminated the initial reads of the long up-the-ramp data. To prevent this part of the curve from affecting our non-linearity correction curve, we extrapolated the low flux data points using a constrained $ax^2 + b$ polynomial.

When we subtract the first Pedestal readout we are subtracting arbitrary counts from each pixel. We need to add the real photon counts back into the up-the-ramp data before applying non-linearity correction. The slope image without non-linearity correction is first generated to estimate the flux in the Pedestal frame. It is then added to the up-the-ramp data before applying the non-linearity correction.

Our measurements of reciprocity failure in HPF's H2RG detector did not show any significant effects of reciprocity failure to correct.

5. DYNAMIC THRESHOLDING

The step after non-linearity correction is thresholding of the up-the-ramp data to contain only the reliable linear part of the corrected up-the-ramp curves on a pixel by pixel basis. The up-the-ramp readout mode provides us power to dynamically threshold at pixel by pixel level depending on the epoch of individual pixel's saturation. However, even though it is desirable for certain optimal signal applications, for precision RV, it is important to have same epoch of observation across the whole order of the spectrum. We constrain the saturation thresholding of pixels of same echelle order together with a user defined echelle order mask.

6. COSMIC RAY HIT RECOVERY

Figure 8 shows Cosmic-Ray hits, Reset anomalies and other artifacts typically seen in an H2RG array readout. Up-the-ramp data enables us to detect CR hits by detecting these signals and recover the signal from unaffected parts of the ramp data. We designed a filter system using a median smooth filter + [1,-3,3,-1] digital abrupt change detection filter to detect CR hits as well as reset anomaly pixels. The slopes are calculated from the clean parts of the affected ramp and combined by variance weighted average.

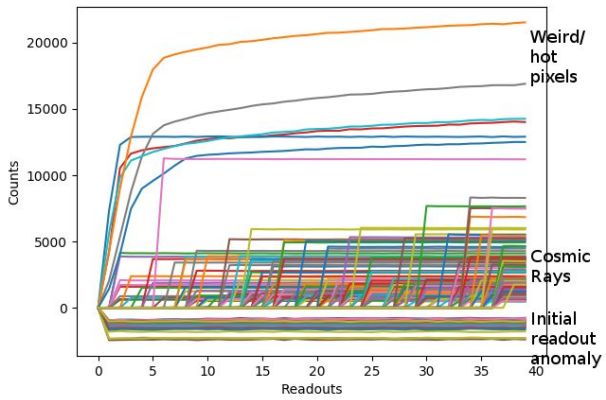

Figure 8. Sample collection of various artifacts typically seen in some pixels in an up-the-ramp data. The Cosmic Ray pixels and initial readout anomaly pixels are recoverable by identifying the exact frames of the events in the up-the-ramp data.

7. FLUX/SLOPE IMAGE CALCULATION

The final slope of up-the-ramp data of each pixel is calculated by the analytical least square linear regression formula

$$b = \frac{nS_{xy} - S_x S_y}{nS_{xx} - S_x^2}\,, \tag{2}$$

where b is the slope, n is the number of readouts, $S_{xy} = \Sigma x_i y_i$, $S_x = \Sigma x_i$, $S_y = \Sigma y_i$, $S_{xx} = \Sigma x_i x_i$, where x is the time vector and y is the counts vector.

The realistic on-sky up-the-ramp data from a star through telescope is never a straight line. Passing clouds, telescope tracking error, jitter, HET's pupil size change, etc., results in a complicated time varying up-the-ramp curve. Usage of this direct analytical formula for slope calculation instead a general least square minimization procedure was chosen for the following robustness properties[‖].

7.1 Robustness of Slope image

7.1.1 Relative flux preservation against throughput changes

To model time variable flux due to clouds, telescope tracking error, HET pupil size changes, etc., in the Equation 2, y_i can be modeled as $y_i = \sum_{j=0}^{j=i} c_j y_T = y_T \sum_{j=0}^{j=i} c_j$, where y_T is the true flux, and c_j is the time varying throughput fraction. Perfect linear data is the special case of $c_j = 1$. Any achromatic throughput changes in the system are represented by these coefficients $C_i = \sum_{j=0}^{j=i} c_j$, enabling us to write $y_i = y_T C_i$.

Consider any two wavelengths ($w1$ & $w2$) in the spectrum which has relative flux difference α, i.e. $y_{T_{w1}} = \alpha y_{T_{w2}}$. Substituting into the equation above, we obtain $S_{xy_{w1}} = \alpha S_{xy_{w2}}$ and $S_{y_{w1}} = \alpha S_{y_{w2}}$., and therefore $b_{w1} = \alpha b_{w2}$. Thus Equation 2 preserves relative flux irrespective of the achromatic[**] throughput changes, c_j.

7.1.2 Associative property of background variations

Errors due to telescope tracking or passing clouds changes the sky background differently from the star flux. Detector bias fluctuations also behave same way. In order to subtract these effects by sky fiber subtraction, the associative property of this fluctuation induced slope changes has to be guaranteed. Using the same formalism as in previous section, Let B_i be the cumulative background variation till the i^{th} readout. Like earlier, we can write $y_i = y_T C_i + B_i$. Then, $S_y = \Sigma y_i = \Sigma y_T C_i + \Sigma B_i = S_{y_T} + S_B$, and $S_{xy} = \Sigma x_i y_i = \Sigma x_i y_T C_i + \Sigma x_i B_i = S_{xy_T} + S_{xB}$. Substituting them into Equation 2, we obtain the slope due to background fluctuation is addictive over the slope of the star flux. i.e. $b = b_T + b_B$. Hence, subtraction of sky fiber which samples the same background can be done on slope images to remove any independently varying background sources like sky or detector bias.

We also verified the robustness of slope images against the above mentioned scenarios by running a Monte-Carlo simulation.

7.2 Variance image and Flux weighted exposure time

For correct error propagation, a per pixel variance image is also calculated along with the slope/flux image using Equation 1 with $\sigma_{bias} = 0$. This variance map is propagated down the HPF pipeline till final RV estimation.

For accurate barycentric velocity correction to obtain the most precision RV, the calculation of the flux weighted midpoint of the observation is important. Up-the-ramp readout of H2RG detector contains information on how the flux varied over an exposure at the cadence of a single frame readout. The cumulative star flux per each order after background and sky subtraction in up-the-ramp data is calculated to obtain the flux weighted exposure time, enabling per-order barycentric correction if needed.

8. SOFTWARE TOOL

All of the H2RG data processing algorithms are written as pip installable Python package `HxRGproc` with support for both Python 2.7 and 3.6+.

To enable maximum modularity and re-usability of the algorithms, `HxRGproc` module is written in functional style. Objects of object oriented programming are used only for data structures or for passing configuration settings. All the main functions are self reliant, and reusable individually. Addition of new features or instrument typically involves adding new functions and passing them as arguments to the existing functions.

[‖] Another important advantage of this simple analytical formula is that it is computationally inexpensive to calculate and it is easily vectorisable.

[**] Chromatic throughput changes will not preserve relative flux, this is an issue even for normal correlated double sample images or for CCD images.

To optimally utilize modern multi-core CPUs and clusters. `HxRGproc` uses multiprocessing module of python to parallelize many of the embarrassingly parallel procedures while processing multiple sets of data. Expensive Python loops are avoided and calculations are all done by vectorized algorithms.

It also generates a standardized and comprehensive trace back logs even while running in parallel mode.

The package has built-in unit tests which cover crucial parts of the code. We are working on increasing the test coverage.

`HxRGproc` also includes submodule for simulation of H2RG up-the-ramp data for arbitrary flux images and telescope tracking errors and jitters. It uses noise generator developed by 6 for simulating H2RG detector noise.

The package also includes many handy utility submodules to generate quality reports and other logs.

9. CONCLUSION AND ONGOING WORK

We have presented an end-to-end comprehensive procedure and software to convert up-the-ramp H2RG data to slope/flux images. This is the first step of processing of H2RG data in our HPF pipeline. Our improved algorithms significantly reduces the bias fluctuation induced non-Gaussian errors in the final flux image.

While these algorithms have already enabled us to demonstrate HPF's radial velocity precision requirements, we are developing more higher order bias correction techniques to improve this further. Two other crucial correction models we are developing is for persistence correction of images with non zero flux, and sub-pixel flat model to correct for sub-pixel sensitivity differences along the hatch patterns on the detector. Our intent is to continually improve these algorithms building on modern statistical techniques and modern computing capability.

ACKNOWLEDGMENTS

This work was partially supported by the funding from The Center for Exoplanet and Habitable Worlds. The Center for Exoplanet and Habitable Worlds is supported by The Pennsylvania State University, The Eberly College of Science, and The Pennsylvania Space Grant Consortium. We acknowledge support from NSF grants, AST1006676, AST1126413, AST1310885, and the NASA Astrobiology Institute (NNA09DA76A) in our pursuit of precision radial velocities in NIR, and support from the Heising-Simons Foundation. Computations for this research were performed on the Pennsylvania State Universitys Institute for CyberScience Advanced CyberInfrastructure (ICS-ACI).

REFERENCES

[1] Mahadevan, S., Ramsey, L., Bender, C., Terrien, R., Wright, J. T., Halverson, S., Hearty, F., Nelson, M., Burton, A., Redman, S., Osterman, S., Diddams, S., Kasting, J., Endl, M., and Deshpande, R., "The habitable-zone planet finder: a stabilized fiber-fed NIR spectrograph for the Hobby-Eberly Telescope," 84461S (Sept. 2012).

[2] Mahadevan, S., Ramsey, L. W., Terrien, R., Halverson, S., Roy, A., Hearty, F., Levi, E., Stefansson, G. K., Robertson, P., Bender, C., Schwab, C., and Nelson, M., "The Habitable-zone Planet Finder: A status update on the development of a stabilized fiber-fed near-infrared spectrograph for the for the Hobby-Eberly telescope," 91471G (July 2014).

[3] Stefansson, G., Hearty, F., Robertson, P., Mahadevan, S., Anderson, T., Levi, E., Bender, C., Nelson, M., Monson, A., Blank, B., Halverson, S., Henderson, C., Ramsey, L., Roy, A., Schwab, C., and Terrien, R., "A VERSATILE TECHNIQUE TO ENABLE SUB-MILLI-KELVIN INSTRUMENT STABILITY FOR PRECISE RADIAL VELOCITY MEASUREMENTS: TESTS WITH THE HABITABLE-ZONE PLANET FINDER," *The Astrophysical Journal* **833**, 175 (Dec. 2016).

[4] Blank, R., Anglin, S., Beletic, J. W., Baia, Y., Buck, S., Bhargava, S., Chen, J., Cooper, D., Eads, M., Farris, M., Hall, D. N. B., Hodapp, K. W., Lavelle, W., Loose, M., Luppino, G., Piquette, E., Ricardo, R., Sprafke, T., Starr, B., Xu, M., and Zandian, M., "The HxRG Family of High Performance Image Sensors for Astronomy," in [*Solar Polarization 6*], Kuhn, J. R., Harrington, D. M., Lin, H., Berdyugina, S. V., Trujillo-Bueno, J., Keil, S. L., and Rimmele, T., eds., *Astronomical Society of the Pacific Conference Series* **437**, 383 (Apr. 2011).

[5] Loose, M., Beletic, J., Garnett, J., and Xu, M., "High-performance focal plane arrays based on the HAWAII-2rg/4g and the SIDECAR ASIC," 66900C (Sept. 2007).

[6] Rauscher, B. J., "Teledyne H1rg, H2rg, and H4rg Noise Generator," *Publications of the Astronomical Society of the Pacific* **127**, 1144–1151 (Nov. 2015).

[7] Rauscher, B. J., Arendt, R. G., Fixsen, D. J., Greenhouse, M. A., Lander, M., Lindler, D., Loose, M., Moseley, S. H., Mott, D. B., Wen, Y., Wilson, D. V., and Xenophontos, C., "Improved Reference Sampling and Subtraction: A Technique for Reducing the Read Noise of Near-infrared Detector Systems," *Publications of the Astronomical Society of the Pacific* **129**, 105003 (Oct. 2017).

[8] Ingraham, P., Perrin, M. D., Sadakuni, N., Ruffio, J.-B., Maire, J., Chilcote, J., Larkin, J., Marchis, F., Galicher, R., and Weiss, J., "Gemini planet imager observational calibrations II: detector performance and calibration," 91477O (July 2014).

[9] Dorn, R. J., Eschbaumer, S., Finger, G., Ives, D., Meyer, M., and Stegmeier, J., "SIDECAR ASIC at ESO," 774219 (July 2010).

[10] Kafadar, K., "The Efficiency of the Biweight as a Robust Estimator of Location," *Journal of Research of the National Bureau of Standards* **88**, 105 (Mar. 1983).

[11] Kubik, B., Barbier, R., Castera, A., Chabanat, E., Ferriol, S., and Smadja, G., "Impact of noise covariance and nonlinearities in NIR H2rg detectors," 91541Q (July 2014).

[12] Biesiadzinski, T., Lorenzon, W., Newman, R., Schubnell, M., Tarl, G., and Weaverdyck, C., "Measurement of Reciprocity Failure in Near-Infrared Detectors," *Publications of the Astronomical Society of the Pacific* **123**, 179–186 (Feb. 2011).